Fourier Analytic Approach to Phase Estimation

Hiroshi Imai[1] and Masahito Hayashi[2]

[1]*National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan*
[2]*Graduate School of Information Sciences, Tohoku University, Sendai, 980-8579, Japan*

For a unified analysis on the phase estimation, we focus on the limiting distribution. It is shown that the limiting distribution can be given by the absolute square of the Fourier transform of L^2 function whose support belongs to $[-1, 1]$. Using this relation, we study the relation between the variance of the limiting distribution and its tail probability. As our result, we prove that the protocol minimizing the asymptotic variance does not minimize the tail probability. Depending on the width of interval, we derive the estimation protocol minimizing the tail probability out of a given interval. Such an optimal protocol is given by a prolate spheroidal wave function which often appears in wavelet or time-limited Fourier analysis. Also, the minimum confidence interval is derived with the framework of interval estimation that assures a given confidence coefficient.

I. INTRODUCTION

Estimating/identifying the unknown unitary operator is discussed in both research fields of quantum computation [1, 2] and quantum statistical inference, therefore, it is a fundamental topic in quantum information. In quantum computation, the unknown unitary operator is given as an oracle, and it is discussed how many applications are required for identifying the given unknown oracle with a given precision. In quantum statistical inference, in contrast, many researchers optimize the average fidelity or mean square error between the true unitary and the obtained guess. In the both research sides, we optimize the state inputting the unknown unitary as well as the measurement, and quadratic speedup is reported from the both research sides [3, 4, 5, 6]. However, the both areas discuss the same topic based on the different criterion in this manner, and there is an example such that the quadratic speedup appears depending on the choice of the criterion. Therefore, since the relation between both criteria is not clear, it is required to treat this problem with a common framework.

In the present paper, as the most typical example, we focus on the phase estimation, in which, the quadratic speedup was demonstrated experimentally [7]. Kitaev first treated the phase estimation problem in quantum computation viewpoint [8]. Since it appears in Shor's factorization, it is considered as a fundamental topic in quantum computation as well as physics. In order to treat the quadratic speedup more deeply, we focus on the limiting distribution, which provides the stochastic behaviour of the estimate around the neighborhood of the true parameter. That is, it provides the distribution of the random variable $n(\hat{\theta} - \theta)$ when the estimate, the true parameter, and the number of applications are given as $\hat{\theta}$, θ, and n. While the limiting distribution is a common concept in statistics and was studied in estimation of quantum state [10, 11, 12], it was not studied in estimation of the unknown unitary operator systematically.

The concept of 'limiting distribution' is very useful for

the following four points. First, the variance of the limiting distribution gives the asymptotic first order coefficient of the mean square error. Second, the tail probability of the limiting distribution for a given interval provides the tail probability of the interval when the width of the interval behaves the order $\frac{1}{n}$. Third, using the limiting distribution, we can discuss the phase estimation under the framework of interval estimation, whose meaning is explained later. Forth, using this concept, we can treat the number of applications required for attaining the given accuracy, i.e., the error probability and the error bar in the asymptotic framework. That is, the above four advantages correspond to respective criteria. Therefore, limiting distribution provides a unified framework for these criteria. The first three criteria are familiar in statistics, and the forth criterion is familiar in computer science.

In the present paper, we analyze the limiting distribution in the phase estimation systematically, and show that the limiting distribution is expressed by Fourier transform of a square integrable function on the closed interval $[-1, 1]$, which approximately gives the input state in the asymptotic setting.

In the realistic setting, the optimization of the forth criterion is more appropriate than that of the first criterion, i.e., the optimization of the mean square error or the average fidelity. However, the estimator only with several special input states were treated in the forth formulation, and its optimization was not discussed while the optimization concerning the mean square error and the average fidelity were done in the existing researches [4, 13].

In the statistics, in order to treat this problem, they consider interval estimation, in which, our inference is given as an interval. Indeed, there are two formulations in the statistics; one is the point estimation, in which, our estimate is given as only one point, and the other is the interval estimation. In the point estimation, it is not easy to guarantee the quality of our estimate because the estimated value always has statistical fluctuation. In order to resolve this problem, in the interval estimation,

for given data and confidence coefficient, our estimate is given as a interval, which is called a confidence interval. In this formulation, smaller width of confidence interval is better. As is mentioned in Section II, when number n of data is sufficiently large, the confidence interval can be provided by using limiting distribution [14].

Further, we analyze the variance and the tail probability of limiting distribution. As our result, concerning the limiting distribution, we prove that the limiting distribution minimizing the asymptotic variance does not minimize the tail probability. It is also shown that the limiting distribution minimizing the tail probability depends on the width of interval. The definition of the tail probability depends on the width of interval. Such an optimal input state is given by a prolate spheroidal wave function[15] which often appears in wavelet or time-limited Fourier analysis. This function is a solution of the linear differential equation [16]:

$$\frac{d}{dx}(1-x^2)\frac{df}{dx} + (\xi(R) - R^2 x^2)f = 0.$$

Originally, prolate spheroidal wave functions appears in analysis on Helmholtz equation in electromagnetics[17] or determination of laser mode [18]. Employing this wave function, Slepian and Pollak [15] extended Shannon's sampling theorem the case where the time-interval is limited as well as the bandwidth while the original Shannon's sampling theorem treats the bandwidth-limited case.

Further, by using these facts, we study optimal interval estimation. We provide the estimation protocol minimizing the width of confidence interval that assures the given confidence coefficient. In this case, the optimal estimation protocol depends on a given confidence coefficient. That means, we must prepare input state properly depending on the confidence coefficient.

The paper is organized as follows. In Section II, the formulation of the phase estimation is given and the limiting distribution is introduced with explanation of its meaning. In Section III, we clarify the relation between the limiting distribution and Fourier transform. In Section IV, we analyze the variance of the limiting distribution. This problem is reduced to find the minimum eigenvalue of a operator in the Dirichlet problem. In section V, the tail probability of the limiting distribution is discussed. It is shown that the limiting distribution minimizing the variance does not provide a small tail probability. In section VI, we treat interval estimation problem. This problem can be analyzed by a prolate spheroidal wave function and the eigenvalue of the defining differential equation. In section VII, the phase estimation with a single copy is discussed in the continuous system. The discussions in Sections IV and V can be applied to this formulation under the deterministic energy constraint. In section VIII, we shortly note on the asymptotic Cramér-Rao lower-bound.

II. LIMITING DISTRIBUTION

Let us consider the estimation problem of the unknown phase shift θ with an n-fold unitary evolution $V_\theta^{\otimes n}$ of the unitary $V_\theta := \begin{bmatrix} e^{i\frac{\theta}{2}} & 0 \\ 0 & e^{-i\frac{\theta}{2}} \end{bmatrix}$, in which our estimating protocol is given by a combination of an appropriate input state $|\phi_0\rangle$ and a suitable measurement M (See Fig. 1).

$$|\psi\rangle \rightarrow \boxed{V_\theta^{\otimes n}} \rightarrow |\psi_\theta\rangle \rightarrow \boxed{M} \rightarrow P_{\theta,\psi}^M$$

FIG. 1: Our estimation scheme

As is mentioned later, this formulation essentially contains the most general framework with n applications of the unknown unitary. In the following discussion, the true parameter is described by θ and our estimate is by $\hat{\theta}$.

As is shown in Appendix A, this problem is equivalent to estimate the parameter θ of the unitary operation:

$$U_\theta = \sum_{k=0}^{n} e^{i(k-\frac{n}{2})\theta} |k\rangle \langle k|. \tag{1}$$

Our scheme for estimating θ is as follows. First, prepare an input state $|\phi_0\rangle = \sum_{k=0}^{n} a_k |k\rangle$ where the coefficients $\vec{a} = \{a_k\}_{k=0}^{n}$ satisfy the normalizing condition $\sum_k |a_k|^2 = 1$. Second, evolve the input state $|\phi_0\rangle$ by the unitary evolution U_θ. And the last, perform a measurement described by a POVM $M = M(\hat{\theta})d\hat{\theta}$. Then, the estimate $\hat{\theta}$ obeys the probability distribution

$$\mathrm{P}_{\theta,\vec{a}}^M(\hat{\theta}) := \langle\phi_\theta| M(\hat{\theta}) |\phi_\theta\rangle,$$

where

$$|\phi_\theta\rangle := U_\theta |\phi_0\rangle = \sum_{k=0}^{n} a_k e^{i(k-\frac{n}{2})\theta} |k\rangle.$$

When our error function is given by $R(\theta,\hat{\theta})$, we optimize the mean error $D_\theta(M,\vec{a}) := \int_0^{2\pi} R(\theta,\hat{\theta})\mathrm{P}_{\theta,\vec{a}}^M(\hat{\theta})d\hat{\theta}$. We only consider the covariant framework, i.e., the error function $R(\theta,\hat{\theta})$ is assumed to be given by a function of the difference $(\theta - \hat{\theta}) \mathrm{Mod}\, 2\pi\mathbb{Z}$. For example, when we focus on the gate fidelity $|\frac{\mathrm{Tr}\, V_\theta V_{\hat{\theta}}^{-1}}{2}|^2 = \cos^2\frac{\theta-\hat{\theta}}{2}$, the error $1 - |\frac{\mathrm{Tr}\, V_\theta V_{\hat{\theta}}^{-1}}{2}|^2 = \sin^2\frac{\theta-\hat{\theta}}{2}$ satisfies the covariant condition.

Then, our measurement may be restricted into a group covariant measurement

$$M_{|t\rangle}(\mathrm{d}\hat{\theta}) := U_{\hat{\theta}}|t\rangle\langle t|U_{\hat{\theta}}^\dagger \frac{\mathrm{d}\hat{\theta}}{2\pi}, \tag{2}$$

where

$$|t\rangle = \sum_{k=0}^{n} e^{i\xi_k} |k\rangle. \tag{3}$$

This is because the minimum of the Bayesian average value $\min_M \frac{1}{2\pi} \int_0^{2\pi} D_\theta(M, \vec{a}) d\theta$ under the invariant prior and the mini-max value $\min_M \max_\theta D_\theta(M, \vec{a})$ can be attained by the same group covariant measurement[19]. Therefore, we restrict our measurement to covariant measurements in the following discussion. Then, our protocol is described by the pair of the coefficient of the input state $\vec{a}$ and the vector $|t\rangle$ given in (3).

Further, without loss of generality, we can restrict our protocol to the pair of $(\vec{a}, |t_0\rangle)$ as follows, where

$$|t_0\rangle = \sum_{k=0}^{n} |k\rangle. \qquad (4)$$

For any protocol $(\vec{a}, |t\rangle)$, we define $\vec{a}' = \{a_k\}$ by

$$a'_k := a_k e^{-i\xi_k}. \qquad (5)$$

Then, as is explained below, the protocol $(\vec{a}', |t_0\rangle)$ has the same performance as the protocol $(\vec{a}, |t_0\rangle)$.

$$
\begin{aligned}
\mathrm{P}_{\theta,\vec{a}}^{M_{|t\rangle}}(\hat{\theta}) &= |\langle \phi_\theta | U_{\hat{\theta}} | t \rangle|^2 = \left| \sum_{k=0}^{n} \overline{a_k} \langle k | U_{\hat{\theta}-\theta} t \rangle \right|^2 \\
&= \left| \sum_{k=0}^{n} \overline{a_k} e^{i\xi_k} e^{i(k-\frac{n}{2})(\hat{\theta}-\theta)} \right|^2 = \left| \sum_{k=0}^{n} \overline{a_k} e^{i\xi_k} e^{i(k-\frac{n}{2})(\hat{\theta}-\theta)} \right|^2 \\
&= \left| \sum_{k=0}^{n} \overline{a'_k} e^{i(k-\frac{n}{2})(\hat{\theta}-\theta)} \right|^2 = \mathrm{P}_{\theta,\vec{a}'}^{M_{|t_0\rangle}}(\hat{\theta}).
\end{aligned}
$$

Therefore, the choice of our protocol is essentially given by the choice of input state. Dam et al [24] proved this argument in a more general framework as follows. When the number of application is n, any protocol can be simulated by the above formulation. That is, any adaptive application of the unknown unitary V_θ can be simulated by the n-fold unitary evolution $V_\theta^{\otimes n}$ under the above error function.

The main target of the present paper is analyzing the asymptotic behavior of output distribution for the sequence of input states $\mathcal{M} := \{\vec{a}^n\}$. For this purpose, we treat the distribution concerning the parameter $z_n = \frac{n(\hat{\theta}_n - \theta)}{2}$ because the estimate $\hat{\theta}_n$ approaches the true parameter θ with the order $\frac{1}{n}$ when an appropriate measurement and an appropriate input state are used. When the random variable z_n converges to a random variable z in probability, the distribution $P(\mathcal{M})$ of z is called the limiting distribution of the sequence of input states $\mathcal{M}$. In the case of state estimation including the classical case, if we apply a suitable estimator, the limiting distribution is the Gaussian distribution under a suitable regularity condition. In the classical case, more precisely, the asymptotic sufficient statistics for the given parameter obeys this Gaussian distribution. That is, any estimator is given as a function of the statistics obeying the Gaussian distribution, asymptotically. This Gaussian distribution is characterized only by the variance.

Even in the quantum case of state estimation, the estimation problem can be reduced in that of quantum Gaussian states family in the asymptotic sense[10, 11, 12]. In particular, if we treat the estimation of one-parameter model, we obtain the same conclusion as is in the classical case. Hence, it is sufficient to evaluate the variance for considering the limiting distribution. That is, there is no variety concerning the limiting distribution of the state estimation because the limiting distribution is uniquely determined as the Gaussian distribution. The main problem in the present paper is, on the other hand, considering whether there exist a variety concerning the limiting distribution of the phase estimation.

When the cost function $R(\theta, \hat{\theta})$ has the form $R(\theta, \hat{\theta}) \cong c(\theta - \hat{\theta})^2 + o((\theta - \hat{\theta})^2)$, the average error behaves as $D_\theta(M_{|t_0\rangle}, \vec{a}_n) \cong \frac{c}{n^2} V$, where V is the variance of the limiting distribution. For example, $R(\theta, \hat{\theta}) = \sin^2 \frac{\theta - \hat{\theta}}{2}$, the constant c is $\frac{1}{4}$. Hence, the analysis on the limiting distribution yields the asymptotic analysis on average gate fidelity. Further, the analysis on the limiting distribution provides the asymptotic analysis on the phase estimation from another aspect. For example, the tail probability of the sequence of input states $\mathcal{M}$ can be described as follows:

$$\mathrm{P}_{\theta,\vec{a}^n}^{M_{|t_0\rangle}} \{|\hat{\theta}_n - \theta| > \frac{A}{n}\} \to P(\mathcal{M})\{|z| > A\}. \qquad (6)$$

Using this relation, we can evaluate the required number of applications of the unknown unitary gate for the given allowable error width B and the given allowable error probability ϵ as follows. First, we choose A by $P(\mathcal{M})\{|z| > A\} = \epsilon$. Next, we choose n by $\frac{A}{n} = B$, i.e., $n = \frac{A}{B}$. Thus, the required number of applications is equal to $\frac{A}{B}$ if we use the sequence of input states $\mathcal{M}$. The above discussions clarify that the analysis on the limiting distribution yields various types of asymptotic analysis on phase estimation.

Hence, in the present paper, for a deeper and unified asymptotic analysis on phase estimation, we analyze the limiting distribution of the sequence of input states $\mathcal{M}$.

III. RELATION WITH SQUARE INTEGRABLE FUNCTIONS

In this section, we give a remarkable relation between limiting distributions and square integrable functions on $[-1, 1]$.

Theorem 1 *For any sequence of input states $\mathcal{M}$ having the limiting distribution, there exists a function $f \in L^2([-1, 1])$ satisfying the normalizing condition $\int_{-\infty}^{\infty} |f(x)|^2 \, dx = 1$ and*

$$P(\mathcal{M}) = \mathrm{P}^f, \qquad (7)$$

where $\mathrm{P}^f(dz) = |F(f)(y)|^2 \, dz$ *and $F(f)$ is the Fourier transform on $L^2(\mathbb{R})$ of f, i.e., $F(f)(y) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{ixy} dx$.*

Conversely, for a function $f \in L^2([-1,1])$ with the normalizing condition there exists a sequence of input states $\mathcal{M}$ satisfying the condition (7).

Due to this relation, we can reduce the analysis on limiting distributions to the analysis on wave functions on the interval $[-1,1]$. That is, our problem results in Fourier analysis on the interval $[-1,1]$.

Proof: As the first step of the proof, we construct a function $f \in L^2([-1,1])$. for a given sequence of input states $\mathcal{M} := \{\vec{a}^n\}$. For this purpose, we define a function $f \in L^2([-1,1])$ by

$$f_n(x) := \overline{a_k}\sqrt{\frac{n}{2}} \tag{8}$$

for $x \in (x_k - \frac{1}{n+1}, x_k + \frac{1}{n+1}]$, where $x_k := \frac{2k-n}{n+1}$. In the following, the set of the above L^2 functions is denoted by L_n^2. The parameter $z_n = \frac{n(\hat{\theta}_n - \theta)}{2}$ can be replaced by the parameter $y_n = \frac{(n+1)(\hat{\theta}_n - \theta)}{2}$ because the ratio $\frac{y_n}{z_n} \to 1$. Since

$$\int_{x_k - \frac{1}{n+1}}^{x_k + \frac{1}{n+1}} f_n(x) e^{ixy_n} dx$$
$$= \overline{a_k^n}\sqrt{\frac{n+1}{2}}\frac{2}{y_n}\sin\frac{y_n}{n+1},$$

we have

$$P_{\theta,\vec{a}^n}^{M_{|t_0\rangle}}(\hat{\theta}_n)\frac{1}{2\pi}\frac{d\hat{\theta}_n}{dy_n}dy_n$$
$$= \frac{1}{\pi(n+1)}\left|\sum_{k=0}^n \overline{a_k^n}e^{i(k-\frac{n}{2})(\hat{\theta}_n-\theta)}\right|^2 dy_n$$
$$= \frac{1}{2\pi}\left|\sum_{k=0}^n \sqrt{\frac{2}{n+1}}\overline{a_k^n}e^{ix_k y_n}\right|^2 dy_n$$
$$= \frac{1}{2\pi}\left|\int_{-1}^1 f_n(x)e^{ixy_n}dx\frac{\frac{y_n}{n+1}}{\sin\frac{y_n}{n+1}}\right|^2 dy_n. \tag{9}$$

When f_n goes to a function $f \in L^2([-1,1])$, since

$$\frac{\frac{y}{n+1}}{\sin\frac{y}{n+1}} \to 1 \tag{10}$$

as n goes to infinity, the distribution of the normalized outcome $y = \frac{(n+1)(\hat{\theta}_n-\theta)}{2}(= y_n)$ convergences to the distribution $P^f(dy)$. If f_n does not converge, by replacing the sequence $\{f_n\}$ by a converging subsequence $\{f_{n_k}\}$, i.e., $f_{n_k} \to f$, we can show that the distribution of the normalized outcome converges the distribution $P^f(dy)$.

Next, we prove the opposite argument. For a given function $f \in L^2([-1,1])$ satisfying the normalizing condition, we construct a sequence of input states satisfying (7). There exists a sequence of functions $f_n \in L_n^2$ such that $f_n \to f$. Then, we define the coefficient $a_k^{(n)}$ by

$f_n(x_k) = \overline{a_k^{(n)}}\sqrt{\frac{n}{2}}$. Due to (9) and (10), the sequence of input states $\mathcal{M}_f := \{\vec{a}^{(n)}\}$ satisfies (7). ∎

For example, when the input state is $\sum_{k=0}^n \sqrt{\frac{1}{n+1}}|k\rangle$, the function f is the constant $\sqrt{\frac{1}{2}}$. Since its Fourier transformation is given by $\frac{1}{\sqrt{2\pi}}\frac{\sin y}{y}$, the limiting distribution is described as

$$P^f(dy) = \left(\frac{\sin y}{y}\right)^2\frac{dy}{2\pi}.$$

IV. VARIANCE OF LIMITING DISTRIBUTION

In the previous section, we have shown that limiting distributions P^f of outcomes are acquired through Fourier transforms of wave functions $f \in L^2([-1,1])$ they correspond to coefficients $a_k^{(n)}$ of input states. In this section, let us seek the input state minimizing the variance by utilizing this fact. As is mentioned in Section II, optimizing the first-order coefficient of the variance is equivalent with minimizing the variance $V(f) := \int_{-\infty}^\infty y^2 P^f(dy)$ over all functions $f \in L^2([-1,1])$.

Define the multiplication operator Q and the momentum operator $P = -i\frac{d}{dx}$ on $L^2(\mathbb{R})$. Then, a function $f \in L^2([-1,1])$ satisfies that

$$V(f) = \langle f| F^\dagger Q^2 F |f\rangle = \langle f| P^2 |f\rangle.$$

Under the natural embedding from $L^2([-1,1])$ to $L^2(\mathbb{R})$, the minimum value of $V(f)$ is given by

$$\min_{f\in L^2([-1,1])} \langle f| P^2 |f\rangle,$$

which is nothing but the Dirichlet problem because the restriction of the operator P^2 on $L^2([-1,1])$ is equivalent with the square of the operator $-i\frac{d}{dx}$ on

$$\{f \in L^2([-1,1]) \cap C^1([-1,1])|f(-1) = f(1) = 0\}.$$

That is, the problem is thus reduced to find the minimum eigenvalue of the operator P^2. Its eigenvalues are $\pi^2 m^2 (m = 1,2,\dots)$ and corresponding eigenfunctions are $\phi_m(x) = 2\sqrt{2}\sin\pi m(\frac{x+1}{2})/C_m$ [20] where C_m is a normalizing constant.

Note that a careful treatment is required for a function $f \in L^2([-1,1])$ when the Dirichlet condition $f(-1) = f(1) = 0$ does not hold. In this case, the function f has a discontinuity at 1 or -1 as an element of $L^2(\mathbb{R})$. Hence, the variance $\langle f|P^2|f\rangle$ is infinity. For example, in the case of $f = \sqrt{\frac{1}{2}}$, the variance $\int_{-\infty}^\infty \sin^2 y\frac{dy}{2\pi}$ diverges. In this case, the limiting distribution is obtained with the order $\frac{1}{n^2}$ [1] while the mean square error goes to zero only the order $\frac{1}{n}$, i.e., we have a quadratic speedup concerning limiting distribution, but no quadratic speedup concerning mean square error. This fact is closely related with the divergence of the integral $\int_{-\infty}^\infty \sin^2 y\frac{dy}{2\pi}$.

V. TAIL PROBABILITY OF LIMITING DISTRIBUTION

In the asymptotic statistics, the behavior of tail probability of the limiting distribution is one of the most important properties [21] because it provides the performances of interval estimation and the powers of the one(or two)-side test. Thus, we consider the tail probability of limiting distribution P^{ϕ_m}.

In the statistics, the tail probability of the limiting distribution is often discussed [21]. So, we consider the tail probability of limiting distribution P^{ϕ_m}. In the i.i.d case, the minimum tail probability and minimum variance among limiting distributions can be realized by the same Gaussian distribution. However, in our setting, the Gaussian distribution does not minimize the variance. Hence, it is not clear whether the minimum tail probability can be attained by the same distribution as the minimizing the variance.

Corresponding limiting distributions are acquired through the Fourier transforms:

$$P^{\phi_m}(y) = |F(\phi_m)(y)|^2.$$

Since

$$
\begin{aligned}
&\int_{-1}^{1} C_m \phi_m(x) e^{ixy} \mathrm{d}x \\
&= \frac{1}{\sqrt{2}i} \int_0^1 e^{i(y+m\pi)x} - e^{i(y-m\pi)x} \mathrm{d}x \\
&= \frac{1}{\sqrt{2}i} \left\{ \left[\frac{e^{i(y+m\pi)x}}{i(y+m\pi)} \right]_{x=0}^1 - \left[\frac{e^{i(y-m\pi)x}}{i(y-m\pi)} \right]_{x=0}^1 \right\} \\
&= \frac{1}{\sqrt{2}} (1 - (-1)^m e^{iy}) \left(\frac{1}{y+m\pi} - \frac{1}{y-m\pi} \right) \\
&= \frac{1}{\sqrt{2}} (-1 + (-1)^m e^{iy}) \frac{2m\pi}{y^2 - m^2\pi^2},
\end{aligned}
$$

limiting distributions are

$$P^{\phi_m}(y) = |F(\phi_m)(y)|^2 = \frac{2m^2\pi(1 - (-1)^m \cos y)}{(y^2 - m^2\pi^2)^2 C_m}.$$

Thus, the tail probability of P^{ϕ_m} decreases with the order $O(y^{-4})$. In order to improve the tail probability, we focus on the well-known fact that the Fourier transform of a rapidly decreasing function is also a rapidly decreasing function. In our problem, the support of the original wave function f is included in $[-1, 1]$. Under this condition, f is a rapidly decreasing function if and only if f is smooth function. Note that a rapidly decreasing function does not decrease 'suddenly'. That is, the smoothness is an essential requirement. For example, the function ϕ_m is not smooth at -1 and 1. In the following, we construct a rapidly decreasing wave function f whose support is included in $[-1, 1]$. In this construction, the smoothing at -1 and 1 is essential.

First, functions g_0, g_1, and g_2 are defined by

$$
g_0(x) := \begin{cases} \frac{2\exp(-1/x)}{\sqrt{x}} & \text{if } x > 0 \\ 0 & \text{otherwise}, \end{cases}
$$
$$g_1(x) := g(x+1),$$
$$g_2(x) := g(1-x).$$

Using these functions, we define a rapidly decreasing g_3 whose support is included in $[-1, 1]$ by

$$g_3(x) = g_1(x)g_2(x)/C$$

where C is the normalizing constant. As is checked numerically (See Fig. 4.), this function improves the tail probability.

Now, we analyze the decreasing speed on the tail probability of $P^{g_3}([-R, R]^c)$. Their Fourier transformations are

$$
\begin{aligned}
F(g_0)(y) &= \frac{1}{\sqrt{2}} \frac{\exp(-\sqrt{2|y|})}{\sqrt{|y|}} \exp \operatorname{sgn}(y) i (\sqrt{2|y|} + \frac{\pi}{4}), \\
F(g_1)(y) &= e^{-iy} F(g_0)(y), \\
F(g_2)(y) &= -e^{-iy} F(g_0)(-y), \\
F(g_3)(y) &= \frac{1}{C} \sqrt{2\pi} F(g_1) * F(g_2)(y),
\end{aligned}
$$

where $F(g_1) * F(g_2)$ is the convolution of $F(g_1)$ and $F(g_2)$. When y is sufficiently large, $|F(g_i)(y)|^2 \cong e^{-2\sqrt{2|y|}}$ for $i = 1, 2$, i.e.,

$$\lim_{y \to \infty} \frac{-1}{\sqrt{|y|}} \log |F(g_i)(y)|^2 = 2\sqrt{2}, \quad i = 1, 2.$$

Then, as is shown in Appendix B, we obtain

$$\lim_{y \to \infty} \frac{-1}{\sqrt{|y|}} \log 2\pi |F(g_1) * F(g_2)(y)|^2 \geq 2\sqrt{2}. \tag{11}$$

Therefore, there exists a function f such that the tail probability of P^f is exponentially small and the support is included in $[-1, 1]$. Note that, the above wave function g_3 does not minimize the variance $V(f)$. This fact tells us that the input state minimizing the variance is not optimal concerning the tail probability of the limiting distribution. That is, the optimal input state depends on the choice of the criterion.

Next, we consider the maximization of the probability $P^f([-R, R])$. For this purpose, we denote the natural projection Π_R from $L^2(\mathbb{R})$ to $L^2([-R, R])$. By using the operator $F_R := F^\dagger \Pi_R F$, this probability has the form

$$\langle f|F_R|f \rangle.$$

That is, our aim is the following maximization:

$$
\begin{aligned}
\max_{f \in L^2([-1,1])} \langle f|F_R|f \rangle &= \max_{g \in L^2(\mathbb{R})} \frac{\langle g|\Pi_1 F_R \Pi_1|g \rangle}{\|g\|^2} \\
&= \max_{g \in L^2(\mathbb{R})} \frac{\|\Pi_R F \Pi_1 g\|^2}{\|g\|^2}.
\end{aligned}
$$

This problem is equivalent with the calculation of the maximum eigenvalue of $\Pi_1 F_R \Pi_1$.

Slepian and Pollak [15] showed that the eigenfunction ψ_R of $\Pi_1 F_R \Pi_1$ associated with the maximum eigenvalue is given as a solution of the linear differential equation, which is called prolate spheroidal wave function:

$$\frac{d}{dx}(1-x^2)\frac{df}{dx} + (\xi(R) - R^2 x^2)f = 0,$$

where $\xi(R)$ is chosen depending on the minimum eigenvalue.

Slepian [16] showed that the maximum eigenvalue $\lambda(R)$ of $\Pi_1 F_R \Pi_1$ behaves as

$$1 - \lambda(R) \cong 4\sqrt{\pi R}e^{-2R}(1 - \frac{3}{32R} + O(R^{-2})) \qquad (12)$$

when R is sufficiently large. The numerical calculation of this minimum probability $\min_{f \in L^2([-1,1])} P^f([-R,R]^c)$ is given in Fig. 4. Thus, the minimum probability $\min_{f \in L^2([-1,1])} P^f([-R,R]^c)$ can be evaluated as

$$\lim_{R \to \infty} \frac{-1}{R} \log \min_{f \in L^2([-1,1])} P^f([-R,R]^c) = 2.$$

That is, the minimum tail probability $\min_{f \in L^2([-1,1])} P^f([-R,R]^c)$ goes to zero with exponential rate 2. This optimal value is attained when the input state is given by the eigenfunction ψ_R of $P_1 F_R \Pi_1$ associated with the maximum eigenvalue $\lambda(R)$.

Now, we numerically compare the functions ϕ_1, g_3, ψ_2, and ψ_{10}. The density functions of the distributions P^{ϕ_1}, P^{g_3}, P^{ψ_2}, $P^{\psi_{10}}$ are plotted in Fig. 3. Their tail probabilities are plotted in Fig. 4. The tail probabilities $P^{\psi_2}([-y,y]^c)$ (thick dashed) and $P^{\psi_{10}}([-y,y]^c)$ (thick solid) attain the minimum tail probability $\min_{f \in L^2([-1,1])} P^f([-y,y]^c)$ only at 2 and 10, respectively. The distributions P^{ϕ_1} and P^{ψ_2} concentrate in the range $[-2,2]$, however, their tail probabilities are not decreasing as rapidly as those of the distributions P^{g_3} and $P^{\psi_{10}}$. This comparison indicates that the optimizations of the concentration and the tail probability are not compatible. That is, the distributions of the Fourier transforms of the functions g_3 and ψ_{10} have a small tail probability (Fig. 3). These functions are smooth at -1 and 1. That means, we have checked that the smoothness is closely related to the tail probability.

Next, we generalize this problem slightly, i.e., we maximize the probability $P^f([R_1, R_2])$. In this case, the maximum value coincides with that of $P^f([(R_1-R_2)/2, (R_2-R_1)/2])$, and its maximum is attained by the function $e^{ix(R_1+R_2)/2}\psi_{(R_2-R_1)/2}(x)$.

Since the function $R \mapsto P^f([-R,R])$ is a strictly monotone increasing function, the inverse function $\beta \mapsto R(\beta)$ is a strictly monotone increasing function. Thus,

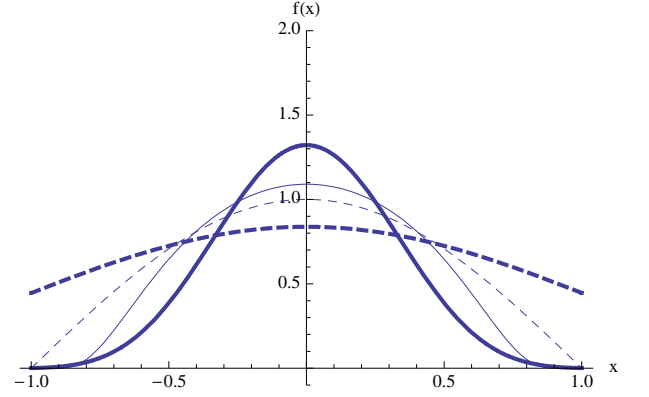$$\min_{f \in L^2([-1,1])} \min\{R | P^f([-R,R]) \ge \beta\} = R(\beta).$$

FIG. 2: The wave function minimizing the variance ϕ_1 (thin dashed), the rapidly decreasing function g_3 whose support is included in $[-1,1]$ (thin solid), prolate spheroidal wave function ψ_2 (thick dashed), and prolate spheroidal wave function ψ_{10} (thick solid).
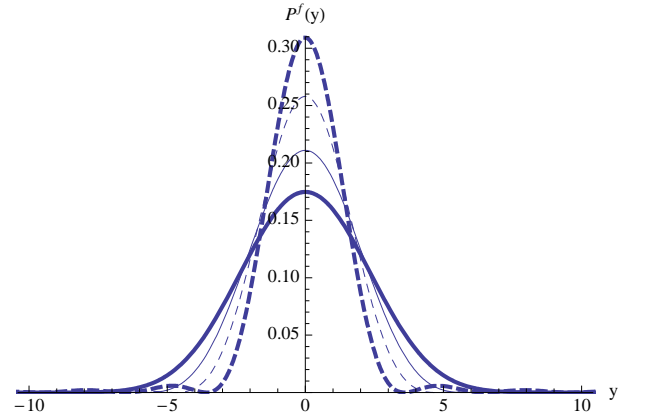

FIG. 3: Probability density function minimizing variance P^{ϕ_1} (thin dashed), probability density function improving tail probability P^{g_3} (thin solid), probability density function P^{ψ_2} (thick dashed), and probability density function $P^{\psi_{10}}$ (thick solid).

Further, the LHS coincides with

$$\min_{f \in L^2([-1,1])} \min\{R | P^f([-R+a, R+a]) \ge \beta\}$$

for any real number a.

VI. INTERVAL ESTIMATION

Now, we treat the phase estimation problem with the interval estimation. In the interval estimation, given a confidence coefficient β, we estimate the confidence interval $[L, U]$, which the unknown parameter θ is guaranteed to belong to with the probability β. Here, since our parameter space is the torus $\mathbb{R}/2\pi\mathbb{Z}$, a careful treatment
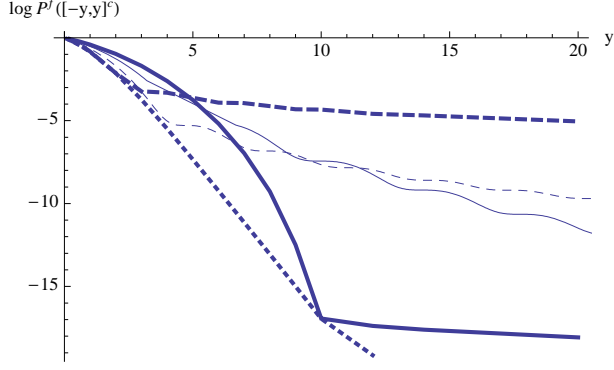
FIG. 4: Logarithm of tail probabilities $\log \mathrm{P}^{\phi_1}([-y,y]^c)$ (thin dashed), $\log \mathrm{P}^{g_3}([-y,y]^c)$ (thin solid), $\log \mathrm{P}^{\psi_2}([-y,y]^c)$ (thick dashed), $\log \mathrm{P}^{\psi_{10}}([-y,y]^c)$ (thick solid), and Logarithm of the minimum tail probability $\log \min_{f \in L^2([-1,1])} \mathrm{P}^f([-y,y]^c)$ (thick dotted).

is required for the confidence interval $[L,U]$. That is, for $L,U \in [0,2\pi)$, the confidence interval $[L,U]$ is defined as a subset of $\mathbb{R}/2\pi\mathbb{Z}$ by

$$[L,U] := \begin{cases} [L,U] & \text{if } L < U \\ [L,U+2\pi] & \text{otherwise,} \end{cases}$$

and its width is defined

$$|[L,U]| := \begin{cases} U - L & \text{if } L < U \\ U + 2\pi - L & \text{otherwise.} \end{cases}$$

In the interval estimation, the upper bound U and the lower bound L of the interval are chosen from the outcome ω obeying the distribution $\mathrm{P}^M_{\theta,\vec{a}}$. Since a smaller width $|[L(\omega),U(\omega)]|$ is better, we minimize the width $|[L(\omega),U(\omega)]|$ with the condition $\mathrm{P}^M_{\theta,\vec{a}}\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta$ for any $\theta \in [0,2\pi)$. That is, we consider

$$\min_{U,L,M,\vec{a}} \left\{ \max_\omega |[L(\omega),U(\omega)]| \Big| \mathrm{P}^M_{\theta,\vec{a}}\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta, \quad \forall \theta \in [0,2\pi) \right\}$$
$$= \min_{U,L,M,\vec{a}} \max_{\theta \in [0,2\pi)} \left(\max_\omega \left\{ |[L(\omega),U(\omega)]| \,\big|\, \mathrm{P}^M_{\theta,\vec{a}}\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta \right\} \right) \tag{13}$$

The value (13) has a mini-max form of the cost $\max_\omega \left\{ |[L(\omega),U(\omega)]| \,\big|\, \mathrm{P}^M_{\theta,\vec{a}}\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta \right\}$, which has a covariant form. Thus, we can restrict our measurement into covariant measurement (2). Hence, our problem is reduced as

$$\min_{t,\vec{a}} \left(\max_\omega \left\{ |[L(\omega),U(\omega)]| \,\big|\, \mathrm{P}^M_{\theta,\vec{a}}\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta \right\} \right).$$

However, since it is quite difficult to treat this optimization with a finite n, we treat the following asymptotic setting as follows:

$$\lim_{n \to \infty} n \min_{t,\vec{a}} \left(\max_\omega \left\{ |[L(\omega),U(\omega)]| \,\big|\, \mathrm{P}^M_{\theta,\vec{a}}\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta \right\} \right)$$
$$= \min_{f \in L^2([-1,1])} \max_\omega \left\{ |[L(\omega),U(\omega)]| \,\big|\, \mathrm{P}^f\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta \right\}$$
$$= \min_{f \in L^2([-1,1])} \min\{2R|\mathrm{P}^f([-R,R]) \geq \beta\}$$
$$= 2R(\beta).$$

This optimal value is attained when the input state constructed by the wave function $\psi_{R(\beta)}$ and the measurement is given by the covariant measurement (2) with the vector $|t_0\rangle$. That is, there exists a pair of functions U and L such that $|[L(\omega),U(\omega)]| \leq 2R(\beta)$ and $\mathrm{P}^M_{\theta,\vec{a}}\{\omega|\theta \in [L(\omega),U(\omega)]\} \geq \beta$. The optimal input state depends on the choice of the confidence coefficient β.

VII. CONTINUOUS CASE WITH SINGLE COPY

Let us consider the phase estimation in the continuous case with single copy, in which by inputing the wave function f, we estimate the parameter θ in a group-covariant model $\rho_\theta = e^{i\theta Q} |f\rangle \langle f| e^{-i\theta Q}$ on the space $L^2(\mathbb{R})$.

It is known that when the shift-covariance condition is assumed for estimators, our estimator is restricted into

the measurement of the observable P[23]. Then, the outcome $\hat{\theta}$ obeys the distribution P^f, and the variance of the outcome is given by $\langle f|\Delta P^2|f\rangle$, which is abbreviated by $\langle\Delta P^2\rangle$.

If we can input any wave function f, the variance can be reduced infinitesimally. Hence, it is natural to assume a constraint for input wave function f. Here, we assume that the potential is given as a monotone function of the absolute value $|Q|$. While we often assume a constraint for average potential, we consider a deterministic condition for potential. That is, the wave packet of f is assumed to exist only in the region where the potential is less than a given constant. In the following, for a simplicity for our analysis, we assume that the input wave function belongs to $L^2([-1,1])$. Hence, the discussion in Sections IV and V can be applied to this problem.

Here, it is meaningful to consider the relation with the Cramér-Rao bound. It is known in general that the Fisher information J_θ for a group-covariant model $\rho_\theta = e^{i\theta Q}|f\rangle\langle f|e^{-i\theta Q}$ is given by $J_\theta = \langle\Delta Q^2\rangle$ because the symmetric logarithmic derivative (SLD) is given by $Q - \langle Q\rangle$[23].

Since the operator P has a commutation relation with Q, we have the Heisenberg limit $\langle\Delta P^2\rangle\langle\Delta Q^2\rangle \geq 1/4$, which is equivalent with the Cramér-Rao inequality:

$$\langle\Delta P^2\rangle \geq \frac{1}{4}J_\theta^{-1}.$$

Especially, if and only if f is a squeezed state satisfying $\langle\Delta P^2\rangle = c$ and $\langle\Delta Q^2\rangle = c^{-1}$, the above inequality is achievable because its attainability is equivalent with that of $\langle\Delta P^2\rangle\langle\Delta Q^2\rangle \geq 1/4$. Thus, if f is not a squeezed state, the Cramér-Rao lower bound $\frac{1}{4}J_\theta^{-1}$ cannot be attained uniformly in the one-copy case. As is shown in the next section, our asymptotic case is essentially equivalent to the above group-covariant model under the restriction of $\mathrm{supp} f \subset [-1,1]$.

VIII. ASYMPTOTIC CRAMÉR-RAO LOWER BOUND

Now, we consider the relation of our discussion with the Cramér-Rao lower bound because the Cramér-Rao approach is often employed in the asymptotic estimation of the unknown unitary[6]. When we apply the sequence of protocols $\mathcal{M} := \{\vec{a}^n\}$, the phase estimation can be treated as the estimation problem in the state family $\{|\phi_{\theta,n}\rangle\langle\phi_{\theta,n}||\theta \in [0,2\pi]\}$, where $|\phi_{\theta,n}\rangle = \sum_{k=0}^n a_k^n e^{ik\theta}|k\rangle$. Let us calculate the SLD Fisher information. From the group covariance of the output state, it suffices to calculate the SLD Fisher information at $\theta = 0$. Let $|l_{\theta,n}\rangle := (1 - |\phi_{\theta,n}\rangle\langle\phi_{\theta,n}|)\frac{\partial}{\partial\theta}|\phi_{\theta,n}\rangle$. The SLD Fisher

information $J_{0,n}$ is given by $\frac{1}{4}J_{0,n} = \langle l_{0,n}|l_{0,n}\rangle$.

$$
\begin{aligned}
\frac{J_{0,n}}{4} &= \langle l_{0,n}|l_{0,n}\rangle \\
&= \langle\phi'_{0,n}|\phi'_{0,n}\rangle - |\langle\phi_{0,n}|\phi'_{0,n}\rangle|^2 \\
&= \sum_{k=0}^n k^2|a_k^n|^2 - \left(\sum_{k=0}^n k|a_k^n|^2\right)^2
\end{aligned}
$$

where $|\phi'_{0,n}\rangle = \frac{\partial}{\partial\theta}|\phi_{\theta,n}\rangle|_{\theta=0}$. Choosing a smooth function f_n by (8), we have

$$\frac{J_{0,n}}{4(n+1)^2} = \sum_{k=0}^n (x_k)^2|f_n(x_k)|^2 - \left(\sum_{k=0}^n (x_k)|f_n(x_k)|^2\right)^2.$$

When f_n converges to f,

$$
\begin{aligned}
\lim_{n\to\infty}\frac{J_{0,n}}{4(n+1)^2} &= \int_{-1}^1 x^2|f(x)|^2\,\mathrm{d}x - \left(\int_{-1}^1 x|f(x)|^2\,\mathrm{d}x\right)^2 \\
&= \langle f|\Delta Q^2|f\rangle.
\end{aligned}
$$

Since the variance of the limiting distribution is $\langle f|\Delta P^2|f\rangle$, we obtain the limiting distribution version of the Cramér-Rao inequality as

$$\langle f|\Delta P^2|f\rangle \geq \frac{1}{\lim_{n\to\infty}\frac{1}{4(n+1)^2}J_{0,n}} = \frac{1}{\langle f|\Delta Q^2|f\rangle}.$$

The equality holds if and only if the wave function f is a squeezed state. However, since the support f belongs to $[-1,1]$, the equality of the above cannot be attained. This fact indicates that the Cramer-Rao approach does not yield the attainable bound in the estimation of unitary action even in the asymptotic formulation, while this approach generally yields the attainable bound in the estimation of quantum state. This point is the essential difference between the state estimation and the unitary estimation.

IX. CONCLUSION

As a unified approach to the asymptotic analysis on the phase estimation, we have treated the limiting distribution on the sequence of estimators because we can recover various asymptotic performance of the estimation protocols from the limiting distribution.

As the first step, we have found a one-to-one correspondence between a limiting distribution and a wave function on $L^2([-1,1])$. That is, we have shown that any limiting distribution is given by the absolute square of the Fourier transform of a wave function $f \in L^2([-1,1])$. Due to this correspondence, it is sufficient to optimize the distribution given as the square of the Fourier transform on $L^2([-1,1])$.

As the next step, the minimization of the variance has been treated among the above distributions by treating

the Dirichlet problem in the similar way as Buzek et al [4]. We have also considered its tail probability. In order to guarantee the small error probability out of the given interval, the limiting distribution is better to be rapidly decreasing. However, it has been clarified that the limiting distribution minimizing the variance is not rapidly decreasing. In order to construct such a limiting distribution, we employ a smoothing method so that we construct a rapidly decreasing function whose support is included in $[-1, 1]$. It has been numerically checked that this function improves the tail probability remarkably.

Further, the tail probability for a given interval has been minimized among these limiting distribution by employing the Slepian and Pollak's analysis on signal processing[15]. The optimal limiting distribution depends on the width of this interval. Using this optimization, we have treat the interval estimation in the asymptotic setting.

Next, we have treated the relation with the phase estimation in the continuous system with the one copy setting. In this case, the Heisenberg's uncertainly relation is equivalent with Cramér-Rao inequality. Using this relation, we have obtained the condition for attainability of Cramér-Rao inequality. Further, we have applied this relation to the asymptotic analysis on the variance of the phase estimation. Then, we have clarified that the Cramér-Rao bound cannot be attained in our framework.

Throughout these discussions, it has been clarified that the optimization of asymptotic phase estimation cannot be characterized by a single parameter while this problem can be characterized by the single parameter, i.e., the variance, in the state estimation of a single parameter model with a regularity condition due to the asymptotic normality[10, 11, 12]. This property is the biggest difference from the state estimation.

Indeed, a similar property can be expected in a general unitary estimation. It is a future problem to investigate the limiting distribution in the estimation of unitary operation in a more general case.

APPENDIX A: ELIMINATION OF MULTIPLICITY

The unitary $V_\theta^{\otimes n}$ can be written as the form

$$U_\theta' = \sum_{k=0}^{n} \sum_{j=1}^{m_k} e^{i(k-\frac{n}{2})\theta} |k, j\rangle \langle k, j|$$

with the multiplicity $m_k = \binom{n}{k}$. When the unitary U_θ' acts on the input state $\sum_{k=0}^{n} \sum_{j=1}^{m_k} a_{k,j} |k, j\rangle$, the final state is given by $\sum_{k=0}^{n} \sum_{j=1}^{m_k} e^{i(k-\frac{n}{2})\theta} a_{k,j} |k, j\rangle = \sum_{k=0}^{n} e^{i(k-\frac{n}{2})\theta} (\sum_{j=1}^{m_k} a_{k,j} |k, j\rangle) = \sum_{k=0}^{n} e^{i(k-\frac{n}{2})\theta} a_k |k\rangle)$ where $a_k := \sqrt{\sum_{j=1}^{m_k} |a_{k,j}|^2}$ and $|k\rangle := \frac{1}{a_k} \sum_{j=1}^{m_k} a_{k,j} |k, j\rangle$. Then, the estimation problem of U_θ' can be reduced in that of U_θ given in (1).

APPENDIX B: PROOF OF (11)

Now, we prove (11). Assume that $y > 0$. For a given integer N,s

$$\frac{\sqrt{2\pi}}{C} | \int_{-\infty}^{\infty} F(g_1)(y')F(g_2)(y - y')dy'|$$

$$\leq \sum_{k=1}^{N} \frac{\sqrt{2\pi}}{C} \max_{y' \in [y\frac{k-1}{N}, y\frac{k}{N}]} |F(g_1)(y')| \cdot |F(g_2)(y - y')| \frac{y}{N}$$

$$+ \frac{\sqrt{2\pi}}{C} \int_{-\infty}^{0} |F(g_1)(y')| \cdot |F(g_2)(y - y')|dy'$$

$$+ \frac{\sqrt{2\pi}}{C} \int_{y}^{\infty} |F(g_1)(y')| \cdot |F(g_2)(y - y')|dy'.$$

Since $|F(g_1)(y')|$ is bounded,

$$\frac{\sqrt{2\pi}}{C} \int_{-\infty}^{0} |F(g_1)(y')| \cdot |F(g_2)(y - y')|dy'$$

$$\cong O(\int_{-\infty}^{0} |F(g_2)(y - y')|dy') \cong O(e^{-\sqrt{2}\sqrt{y}}).$$

Similarly,

$$\frac{\sqrt{2\pi}}{C} \int_{y}^{\infty} |F(g_1)(y')| \cdot |F(g_2)(y - y')|dy' \cong O(e^{-\sqrt{2}\sqrt{y}}).$$

Further,

$$\frac{\sqrt{2\pi}}{C} \max_{y' \in [y\frac{k-1}{N}, y\frac{k}{N}]} |F(g_1)(y')| \cdot |F(g_2)(y - y')| \frac{y}{N}$$

$$\cong O(e^{-\sqrt{2}\sqrt{y\frac{k-1}{N}}} \cdot e^{-\sqrt{2}\sqrt{y - y\frac{k}{N}}}) \leq O(e^{-\sqrt{2}\sqrt{y(1-\frac{1}{N})}}).$$

Therefore,

$$\frac{\sqrt{2\pi}}{C} | \int_{-\infty}^{\infty} F(g_1)(y')F(g_2)(y - y')dy'| \leq O(e^{-\sqrt{2}\sqrt{y(1-\frac{1}{N})}}).$$

Since N is arbitrary,

$$\frac{\sqrt{2\pi}}{C}|\int_{-\infty}^{\infty} F(g_1)(y')F(g_2)(y-y')dy'| \leq O(e^{-\sqrt{2}\sqrt{y}}).$$

Taking the square, we obtain (11).

In the case of $y < 0$, we can show (11) by replacing y' by $-y'$.

---

[1] R. Cleve, A. Ekert, C. Macchiavello and M. Mosca, "Quantum Algorithm Revisited," *Proc. R. Soc. London, Ser. A* 454, 339, 1998.

[2] A. Y. Kitaev, A. H. Shen and M. N. Vyalyi, *Classical and Quantum Computation, (Graduate Studies in Mathematics 47),* Americal Mathematical Society, 2002.

[3] V. Giovannetti, S. Lloyd and L. Maccone, "Quantum-enhanced measurements: beating the standard quantum limit," *Science* 306, 1330-1336, 2004.

[4] V. Buzek, R. Derka and S. Massar, "Optimal Quantum Clocks," *Phys. Rev. Lett.* 82 (1999) 2207, quant-ph/9808042.

[5] M. Hayashi, "Parallel Treatment of Estimation of SU(2) and Phase Estimation," *Phys. Lett. A* 354, 183-189, 2006.

[6] H. Imai and A. Fujiwara, "Geometry of optimal estimation scheme for SU(D) channels," *J. Phys. A* 40, 4391-4400, 2007.

[7] B. L. Higgins, D. W. Berry, S. D. Bartlett, H. M. Wiseman and G. J. Pryde, "Entanglement-free Heisenberg-limited phase estimation," *Nature* 450, 393-396, 2007.

[8] A. Y. Kitaev, "Quantum Computations: Algorithms and Error Correction," *Russ. Math. Surv.* 52, 1191-1249, 1997.

[9] M. Hayashi and K. Matsumoto, "Asymptotic performance of optimal state estimation in quantum two level system," *quant-ph/0411073*.

[10] M. Guţă and J. Kahn, "Local asymptotic normality for qubit states," *Phys. Rev. A*, **73**, 052108 (2006).

[11] M. Guţă, B. Janssens, and J. Kahn, "Optimal estimation of qubit states with continuous time measurements," *Commun. Math. Phys.*, **277**, 127-160 (2008).

[12] M. Guţă and A. Jencova, "Local asymptotic normality in quantum statistics," *Commun. Math. Phys.*, **276**, 341-379 (2007).

[13] D. Pope, H. M. Wiseman and N. K. Langford, "Adaptive Phase estimation is more accurate than nonadaptive phase estimation for continuous beams of light," *Phys. Rev. A* A70, 043812-1 - 043812-13 2004.

[14] R. J. Larsen and M. L. Marx, *An Introduction to Mathematical Statistics and Its Applications*, Pearson Education, U.S., 2005.

[15] D. Slepian and H. O. Pollak, "Prolate spheroidal wave functions, Fourier analysis and uncertainty-I," *Bell Syst. Tech. J.*, 40, 43-63, 1961.

[16] D. Slepian, "Some asymptotic expansions for prolate spheroidal functions," *J. Math. Phys.* 44, 99–140, 1965.

[17] L. Li, M. Leong, T. Yeo and Y. Gan, "Electromagnetic radiation from a prolate spheroidal antenna enclosed in a confocal spheroidal radome," *IEEE Trans. on Antenn. and Propa.*, 50, 1525-1533, 2002.

[18] P. Nazmi, P. Kapadia and J. Dowden, "A mathematical model of heat conduction in a prolate spheroidal coordinate system with applications to the theory of welding," *J. Phys. D*, 26, 563-573, 1993.

[19] A. S. Holevo, "Covariant Measurements and Uncertainty Relations," *Rep. Math. Phys.* 16, 385-400, 1979.

[20] E. A. Coddington and N. Levinson, *Theory of differential equations*, McGraw-Hill, New-York, 1955.

[21] F. P. Kelly (ed.), *Probability, Statistics and Optimization: A Tribute to Peter Whittle (Wiley Series in Probability and Statistics)*, John Wiley & Sons Inc., 1994.

[22] A. S. Holevo, "Asymptotic estimation of shift parameter of a quantum state," *quant-ph/0307225*.

[23] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, (North-Holland, Amsterdam, 1982); Originally published in Russian (1980).

[24] W. van Dam, G. M. D'Ariano, A. Ekert, C. Macchiavello, and M. Mosca, "Optimal quantum circuits for general phase estimation," *quant-ph/0609160*.